PL-TR-92-2027
Environmental Research Papers, No.1097

AD-A256 181

CLOUD DATA SET FOR NEURAL NETWORK CLASSIFICATION STUDIES

Rupert S. Hawkins
K. F. Heideman
Ira G. Smotroff

23 January 1992

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

PHILLIPS LABORATORY
DIRECTORATE OF GEOPHYSICS
AIR FORCE SYSTEMS COMMAND
HANSCOM AIR FORCE BASE, MA 01731-5000

"This technical report has been reviewed and is approved for publication"

JOSEPH WILLIAM SNOW, Chief
Satellite Meteorology Branch

ROBERT A. McCLATCHEY, Director
Atmospheric Sciences Division

This document has been reviewed by the ESD Public Affairs Office (ESD/PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organziation, please notify AFGL/DAA, Hanscom AFB, MA 01731-5000. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
23 Jan 1992

3. REPORT TYPE AND DATES COVERED
Scientific Final

4. TITLE AND SUBTITLE
Cloud Data Set for Neural Network Classification Studies

5. FUNDING NUMBERS
FE 62101F
PR 6670
TA 17
WU 12

6. AUTHOR(S)
Hawkins, Rupert S., K.F. Heideman, and Ira G. Smotroff*

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Phillips Laboratory (GPAS)
Hanscom AFB, MA 01731-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
PL-TR-92-2027
ERP, No. 1097

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
* The MITRE Corporation, Bedford, MA 01730

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release: distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

A large set of GOES imagery has been analyzed to generate a cloud data set for neural network classification studies. The classification scheme uses 13 classes of clouds. The cloud data base is large enough to allow cross-validation tests of automated cloud classification techniques. The methods used to acquire the data on cloud type and a brief overview of the planned automated classification experiments are presented. The appendix contains an inventory of the cloud data set.

14. SUBJECT TERMS
Cloud analysis
Satellite imagery
Cloud classification
Cloud typing

15. NUMBER OF PAGES
24

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Contents

Illustrations

Cloud Data Set for Neural Network Classification Studies

1. INTRODUCTION

This document describes activities carried out jointly by the Phillips Laboratory Geophysics Directorate and the MITRE Corporation to construct a cloud data base of sufficient size to enable cross-validation tests of automated cloud classification techniques. The methods used to acquire the cloud type data and a brief overview of the planned automated classification experiments are presented. The appendix contains an inventory of the cloud data set.

2. BACKGROUND

From the earliest days of meteorological satellites, there have been efforts directed toward systematic analysis of images of the earth and its clouds. In recent years, cloud analysis has been undertaken with the use of the computer. Specifically, the goal has been cloud typing, that is, the specification of regions of images in an automated framework.

Received for Publication 22 January 1992

The value of an automated procedure for specifying particular aspects of images would be great. The products of such a process would be self-consistent, and could be produced in real time at little cost. These results would be the basis of a wide range of studies and operational uses.

The Air Force Global Weather Center at Offutt AFB, Nebraska has operated an automated routine known as the Real-Time Nephanalysis (RTNEPH) since the early seventies. Assessments of 11 cloud types are made using satellite imagery. Those types are cumulus, altocumulus, stratocumulus, stratus, altostratus, nimbostratus, cirrus, cirrocumulus, cumulonimbus, clear, and unknown. The cloud typing technique makes use of statistical sample means and variances of satellite brightness data, and is limited in its usefulness when visible and infrared data are not simultaneously available. The RTNEPH system could incorporate improved cloud typing techniques.

Parikh[1] made a statistical study of classification techniques for a four-group cloud model (low, mixed, cirrus, and cumulonimbus) and a three-group model (low, cirrus, cumulonimbus). She showed a preference toward design parameters for the automatic classification of cloud systems.

Harris and Barrett[2] reported a method for the automatic determination of cloud types and cloud amounts. Textural measures such as standard deviations of brightness and vector dispersion of grayshade density values were evaluated.

Garand[3] developed a classification procedure for oceanic cloud patterns. His classification scheme incorporated 20 classes. His results are very promising, although it is not clear that the hand-crafted image features used could be economically scaled to create a usable system.

Hawkins and d'Entremont[4] presented an automated cloud typing algorithm whereby 1 km visible and infrared Defense Meteorological Satellite Program (DMSP) data are classified into one of 18 cloud types, including clear. The cloud typing algorithm first converts a satellite visual image into a one-bit binary image that has pixel values at one of two possible greyshades. Spectral and spatial information from the original image is retained in this process. Once the one-bit image is constructed, it is then analyzed for features that relate to the spatial variations. Cumulative lengths of runs of "black" and "white" are compared to mean cumulative run-lengths for the cloud types. The closest match is chosen as the cloud type.

---

1 Parikh, J. (1977) A Comparative Study of Classification Techniques. *Remote Sensing of the Environment*, **6**:67-81.

2 Harris, Raymond and Barrett, E.C. (1978) Toward an Objective Nephanalysis. *Journ. Appl. Meteor.*, **17**:1258-1266.

3 Garand, Louis (1988) Automated Recognition of Oceanic Cloud Patterns, Part 1: Methodology and Application of Cloud Climatology. *Journ. of Clim.*, **1**:(No. 1) 20-39.

4 Hawkins, Rupert S. and d'Entremont, R.P. (1990) *Automated Cloud Typing Using Satellite Imagery*. Geophysics Laboratory Report GL-TR-90-0326, Hanscom AFB, MA, 6pp. ADA230492

3. CLOUD CLASSIFICATION PROCEDURE

A large set of GOES imagery has been analyzed to give a cloud data set for neural network classification studies. This section describes that data set. A brief explanation of the analysis methods will also be given.

Data saves were made of GOES-EAST imagery during June and July 1991 of the New England area. Visual (0.55-0.75 μm) and infrared (10.5-12.5 μm) images were saved from 1430 GMT to 1900 GMT on the half hour. As can be seen in the data inventory (see Appendix) some data were missing or were otherwise unusable.

The classification scheme for this study was rather extensive compared to many in use. The trend appears to be towards more classes due to the improved discrimination capability of classification techniques. While Garand[3] used a large number of classes, 20, he included 6 classes that are not strictly classes as seen in prior work. These classes describe the morphology of clouds (for example, Cloud Streets, Roll Clouds, and Open Cells) and include a number of mixed classes (such as Altocumulus with Cumulus, Bright Multi-layers with Cirrus/cumulonimbus). We chose not to include morphology classes because morphology is an automatically computed feature of the new classification scheme and we included only one mixed class to avoid introducing unecessary subjectivity into the training data.

The classification scheme used follows the philosophy of Hawkins and d'Entremont[4]. The following thirteen classes are used:

Clear
Cumulus
Small scattered cumulus
Stratocumulus
Stratus
Altocumulus
Altostratus
Thin cirrus
Cirrus
Thin cirrus over cloud
Cumulonimbus
Fog
Haze

With the use of a land-sea map, the thirteen classes are automatically reclassified into the 26 categories of land-clear, sea-clear, land-cumulus, sea-cumulus, etc. After performing the classifications, our assessment is that the scheme is sound and appropriate for the kind of data set we are working with.

Ira Smotroff of the MITRE Corporation devised an interactive computer program for manually classifying the image data sets and storing files of the results. We refer to this as the knowledge acquisition tool. Classifications may be selected while viewing visible, infrared, or a visible/infrared composite image. A mouse is used to sweep squares of a user selectable size (typically 12 pixels) across the image to mark cloud samples. Each sample has an associated color that encodes the cloud type. A menu bar along the edge of the display is used to select the desired cloud type. Mistakes are easily erased and corrected either during the initial session or later in a resumed session. The result of this process is a new image file called a "Pick" file which contains the classification information in the selected pixel locations.

Two consoles were used side by side as shown in Figure 1. Two analysts manned the consoles and worked together to produce consensus classifications. Both analysts could see both displays. The screen on the left showed the current session. The screen on the right showed the previous half hour analysis with its Pick image superimposed. The right console was used to make modifications and additions as needed. We feel that this double analysis scheme is superior to that possible with a single console and analyst.

Figure 2 shows a typical visible image of the sample region. There is about 60 percent land and 40 percent water for the New England area used. Figure 3 shows the same image with our analysis information superimposed by the knowledge acquisition tool. Figure 4 shows the corresponding infrared image with analysis superimposed by the knowledge acquistion tool. Figure 5 shows the composite vis-infrared image with analysis superimposed by the knowledge acquisition tool. In this case, the analysis appears in one color because of system color table limitations.

A summary of the database analysis appears in the appendix.

4. NEURAL NETWORK CLASSIFICATION STUDY

4.1 Overview

Neural networks are appropriate for meteorological classification tasks for a number of reasons. First, their associative properties allow graceful degradation of performance under conditions of ambiguity and noise, thus avoiding the brittle behavior of many standard approaches. Second, they learn to perform tasks that cannot easily be specified analytically. This allows improved performance in perception tasks and cost-effective retargeting of systems to additional domains. Finally, they can be executed in real-time on appropriate hardware. To exploit these properties, this research is developing a general approach to meteorological classification based on neural network data fusion. The system is being applied to cloud type identification from satellite imagery. However, the system can easily be retrained to perform a range of other meteorological identification tasks such as the identification of hurricanes, thunderstorm outflow boundaries, etc.

Figure 1. Analyst Console Configuration

Figure 2. Typical Visible Image

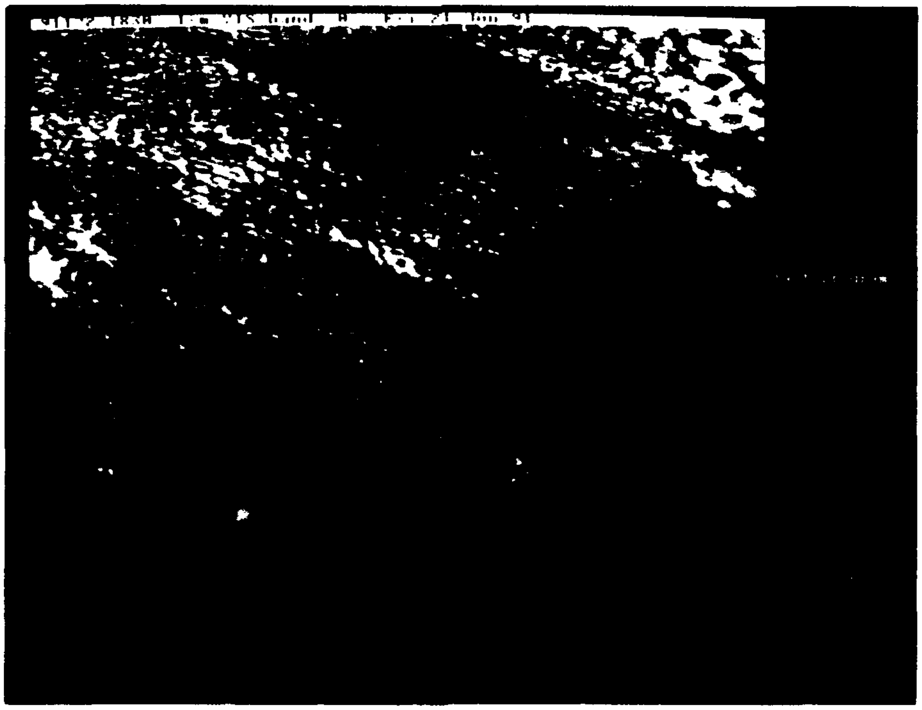

Figure 3. Knowledge Acquisition Tool-Visible Mode

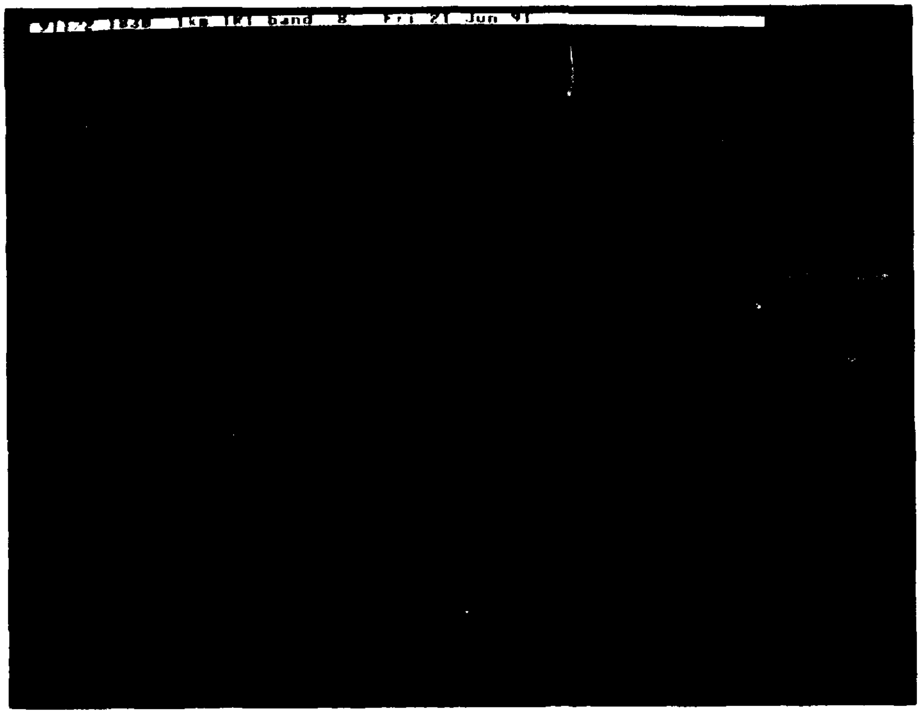

Figure 4. Knowledge Acquisition Tool–Infrared Mode

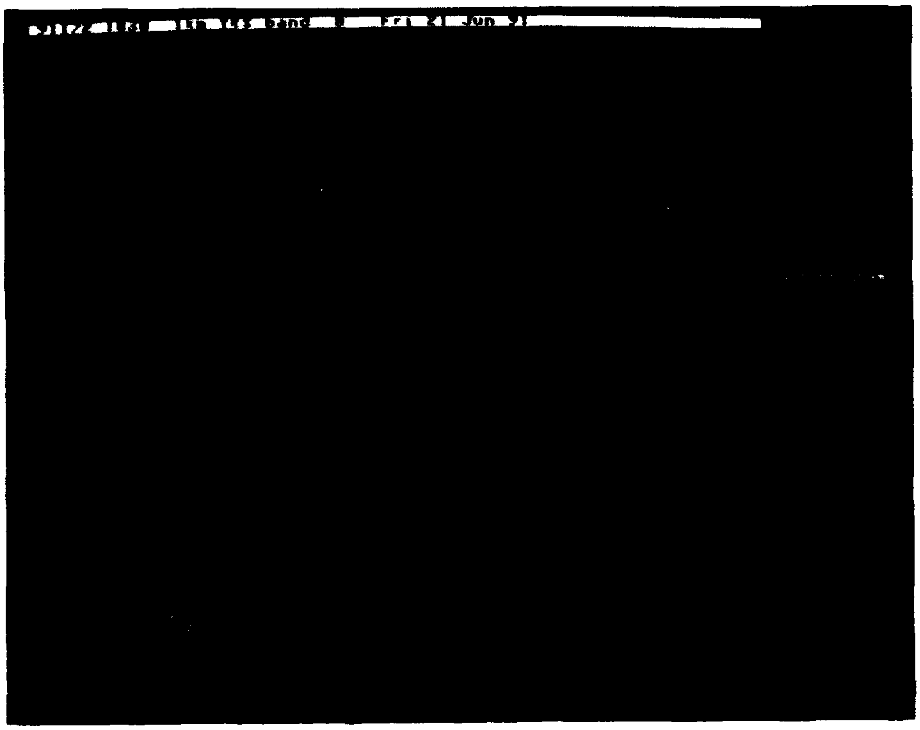

Figure 5. Knowledge Acquisition Tool–Composite Mode

A number of promising preliminary results for the method have been demonstrated during the past 2 years, including highly accurate classification performance on a limited data set and graceful degradation of classification performance over large shifts of terrain. These results point to the applicability of neural networks for automated generation of meteorological products in real time.

The research vehicle is a system called the DATA FUNNEL (DATA Fusion Using Neural NEtwork Learning). The system architecture is shown in Figure 6. Heterogeneous sensor streams including point sensor data and/or image data are fed into the system. A vision system based on a number of neural network theories operating across all image input channels produces a range of non-local products that augment the local training data for the neural network classifier stage. Point sensor data can be optionally extrapolated in two dimensions to match image data. Supervised learning is used to train the classifiers. The aforementioned meteorological database will provide the target signal for training. To allow the inclusion of meteorological heuristics, a knowledge based system controls the performance elements of the classification system. The control component may incorporate both neural network and expert system technologies. Classification performance is measured by cross validation tests using untrained human classifications.

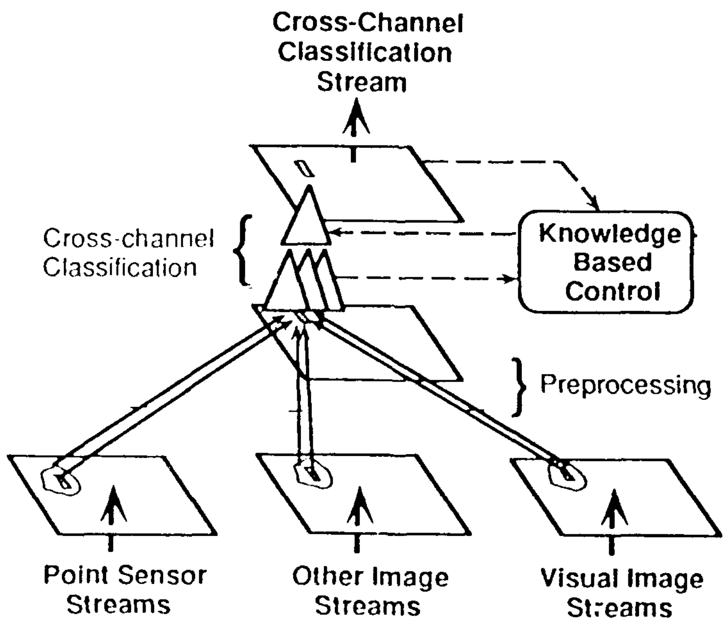

Figure 6: DATA FUNNEL Architecture

Feature vectors for classification are constructed by appending actual data with derived products generated by a number of preprocessing steps. The derived products provide information of a non-local nature which is used as part of a classification that operates on local regions (that is, pixels). Vision algorithms for image data provide texture and morphology features.

The neural network classification sub-system is constructed using self-organizing modular learning algorithms developed by the MITRE Corporation. It implements a number of non-linear discriminant functions specifically tailored to the meteorological task to provide the highest accuracy and generalization to untrained data. The current learning algorithm research is fine-tuning these techniques to improve the ability of our learning algorithms to scale up to the large periods of time and multiple geographic areas that would be required for eventual deployment.

Research is also being conducted to improve the capability of the vision system to reliably segment clouds and to provide morphology information.

4.2 Experimental Goals

The major goal of the experiment is to test the generalization capabilities of the neural network classification system on untrained data. Another goal is to test the effectiveness of various algorithmic improvements. A final goal is to test various strategies for scaling the techniques to operational status.

The cloud database described in Section 2 consists of 10 samples per day taken at half hour intervals. Due to collection problems, a maximum of 27 samples is available for some hours and as few as 22 are available for other hours.

Analyst input marked only the 12 types of clouds and "clear". The MITRE Corporation was provided with three map projections for dates 10 June 91, 25 June 91, and 7 July 91. These were used to construct land/ocean maps for those dates. Using these land/ocean maps, 26 classes were determined from the original 13 classes.

To maximally utilize the data, a series of leave-one-out cross-validation tests will be run, and an average classification accuracy will be computed. These will be run at a single time across the days and at a combination of times to determine the minimal number of networks and training needed to attain good performance on the data set. One useful experiment is to verify the hypothesis that viewing angles equidistant from the zenith will have symmetrical shadows. If the hypothesis is true, they might be combined into a single larger classification data set.

References

1. Parikh, J. (1977) A Comparative Study of Classification Techniques. *Remote Sensing of the Environment*, **6**:67-81.

2. Harris, Raymond and Barrett, E.C. (1978) Toward an Objective Nephanalysis. *Journ. Appl. Meteor.*, **17**:1258-1266.

3. Garand, Louis (1988) Automated Recognition of Oceanic Cloud Patterns, Part 1: Methodology and Application of Cloud Climatology. *Journ. of Clim.*, **1**:(No. 1) 20-39.

4. Hawkins, Rupert S. and d'Entremont, R.P. (1990) *Automated Cloud Typing Using Satellite Imagery.* Geophysics Laboratory Report GL-TR-90-0326, Hanscom AFB, MA, 6pp. ADA230492

Appendix

Summary of the Data Base

	6/3/91	6/4/91	6/5/91	6/6/91	6/7/91
1430	N.A.	N.A.	N.A.	N.A.	N.A.
1500	9,8,5,2,7	5,2,7,8	2,11,7,5	1,2,11,4 8,7	11,2,1
1530	11,8,3,5, 2,7	5,2,7,8	11,2,5,7	1,3,4,2,8, 11	11,2,1
1600	11,4,10,5, 2,7	5,2,7,8 4	11,2,5,7	1,3,4,2,8, 11,5,9	11,2,1
1630	11,5,3,2,4 10	5,2,7,8	11,2,5,7	11,2,3,1,5	11,2,1
1700	11,2,6,5,7 4,3	5,2,7,8	11,2,5,7, 9,8	11,7,2,4,3	11,2,1
1730	11,2,5,4,3, 6	5,2,7,8 9	9,8,5,2 7,11	11,3,4,2,1	11,2,1
1800	11,2,5,3,6	5,2,7,8, 9	11,2,8,9, 5,7	11,3,2,1	11,2,1
1830	11,6,2,5,3 4	3,2,7,8,9	11,2,8,9, 5,7	11,3,5,2,1	11,2,1
1900	11,3,4,5,6	5,2,7,8,9	11,2,8,9, 5,7,3	11,3,5,2,1	11,2,1

Key:
1. small scattered cumulus
2. cumulus
3. thin cirrus
4. cirrus
5. thin cirrus over cloud
6. stratus
7. stratocumulus
8. altocumulus
9. altostratus
10. cumulonimbus
11. clear
12. haze
13. fog

	6/10/91	6/11/91	6/12/91	6/13/91	6/14/91
1430	11,2,4,3 1,7	11,2,4,5	2,7,8,5 11,6	11,8,7,2	N.A.
1500	11,2,4,3 7,1	11,5,4,3, 2	2,7,8,5 11,6	11,8,7,2 1	11,4,3,7, 2
1530	11,2,4,3 7,1	11,5,4,3, 2	2,7,8,5 11,6	11,8,7,2 1	11,4,3,2, 1,7
1600	11,2,4,3, 1,7	11,5,3,2, 4	N.A.	N.A.	11,4,3,2, 1,7
1630	11,2,4,3, 1	11,5,3,2, 4	N.A.	11,2,7,8	11,4,3,2, 1,7
1700	11,2,4,7, 3,1,5	11,5,3,2, 4	11,8,7,2, 6,5	11,2,7,8	11,4,2,3, 1,7
1730	11,2,4,7, 5,1	11,5,3,2, 4	11,8,7,2, 6,5	11,2,7,8	11,4,2,3, 1,7
1800	11,2,4,7, 5,1	11,5,7,2, 4	2,8,11,5, 6	8,7,2,11	11,4,2,3, 1,7
1830	11,2,3,4, 5,7	11,5,7,2, 4,1	2,8,11,5, 10	N.A.	11,4,2,3, 8
1900	11,2,5,3, 7	11,5,7,2, 4,1	N.A.	N.A.	11,4,2,3, 8

Key:

1. small scattered cumulus
2. cumulus
3. thin cirrus
4. cirrus
5. thin cirrus over cloud
6. stratus
7. stratocumulus
8. altocumulus
9. altostratus
10. cumulonimbus
11. clear
12. haze
13. fog

	6/17/91	6/18/91	6/19/91	6/20/91	6/21/91
1430	6,7,5,11	11,2,5,4	11,4,7	2,11,6	11,4,5, 2
1500	6,7,5,11	11,2,6,4	11,4,7	11,2,7,6 1	11,4,5, 2
1530	7,5	N.A.	11,4,7	11,2,7,6 1	11,4,5, 2
1600	7,5	N.A.	11,4,7	11,2,7,6 1	11,4,5, 2
1630	7,5,4,8	N.A.	11,4,7	11,2,6,1	11,4,5, 2
1700	7,2,5,11, 8	N.A.	11,4,7	11,2,6,1, 4	N.A.
1730	5,7,8,11, 2	N.A.	11,4,7	11,2,6,1, 4	11,4,5, 2,3,1
1800	11,7,2	N.A.	11,4,7	N.A.	N.A.
1830	11,7,2,8	N.A.	11,4,7,2	N.A.	11,4,5, 2,3,1
1900	11,7,2,7	N.A.	11,4,7	N.A.	N.A.

Key:

1. small scattered cumulus
2. cumulus
3. thin cirrus
4. cirrus
5. thin cirrus over cloud
6. stratus
7. stratocumulus
8. altocumulus
9. altostratus
10. cumulonimbus
11. clear
12. haze
13. fog

	6/24/91	6/25/91	6/26/91	6/27/91	7/1/91
1430	N.A.	N.A.	N.A.	N.A.	11,4,2, 1
1500	2,1,13,11	2,11,1	1,2,3,4, 5,7,11	11,2,3,5, 4,9,8	11,4,2, 5,1
1530	2,1,13,11	2,11,1	1,2,3,4, 5,7,11	2,11,4,5, 9,8	11,4,2, 5,1
1600	2,1,13,11	2,11,1	1,2,3,4, 5,7,11	2,11,4,1, 8,3	11,4,2, 5,1,3
1630	2,1,13,11	2,11,1	1,2,3,4, 5,7,11	5,11,2,8, 4,3	11,4,2, 5,1,3
1700	2,1,13,11	2,11,1	1,2,3,4, 5,7,11	2,11,4,8, 7,3,5	11,4,2, 5,1,3
1730	2,1,13,11	2,11,1	1,2,3,4, 5,11	2,11,4,8, 1,3,5	11,4,2, 5,1,3
1800	2,1,13,11	2,11,1	1,2,3,4, 5,11	11,3,2,1, 5,8,4	3,2,4, 5
1830	2,1,13,11	2,11,1	1,2,3,4, 5,11	11,2,3,4, 5,8	3,2,4, 5,1
1900	2,1,13,11	2,11,1	1,2,3,4, 5,11	11,2,3,4, 1,5,8	3,2,4, 5,11,1

Key:

1. small scattered cumulus
2. cumulus
3. thin cirrus
4. cirrus
5. thin cirrus over cloud
6. stratus
7. stratocumulus
8. altocumulus
9. altostratus
10. cumulonimbus
11. clear
12. haze
13. fog

	7/1/91	7/2/91	7/8/91	7/10/91	7/11/91
1430	N.A.	11,8,7,2, 9	N.A.	11,2,1,4, 5	11,1,2,8, 6
1500	2,11,7,1, 4,3	11,8,7,2, 9	N.A.	11,2,1	11,1,2,8, 6
1530	11,4,3,5, 7,1	11,8,9,7	11,5,3,8, 7,10	11,2,1	11,1,2,8, 6
1600	2,11,4,3, 5,7,1	11,2,8,9	11,5,3,8, 7,10	11,2,1,7	11,1,2,8, 6,9
1630	2,11,7,1, 4,5	11,2,8,9	11,5,3,8, 7,10	N.A.	N.A.
1700	2,11,7,1, 4,5	N.A.	11,5,3,8, 7,10	N.A.	N.A.
1730	2,11,7,1, 5	N.A.	11,2,4,8, 7,10,1	N.A.	N.A.
1800	2,11,7,1, 4,5	N.A.	11,2,5,4, 8,7,1	N.A.	N.A.
1830	11,7,2,1, 4,5	N.A.	11,2,6,7	N.A.	N.A.
1900	11,7,2,1, 4,3,5	11,2,5,8, 1,7	11,2,6,7, 1	N.A.	N.A.

key:

1. small scattered cumulus
2. cumulus
3. thin cirrus
4. cirrus
5. thin cirrus over cloud
6. stratus
7. stratocumulus
8. altocumulus
9. altostratus
10. cumulonimbus
11. clear
12. haze
13. fog

	7/12/91	7/14/91	7/15/91	7/16/91	7/17/91
1430	11,4,2,1	11,2,6,7	2,11,1	11,2,1	2,7,11,1
1500	11,4,2,5 1	11,2,6,7	2,11,1	11,2,1	2,7,11,1
1530	11,4,2,5, 1	11,2,6,7	2,11,1	11,2,1	2,7,11,1
1600	11,3,2,1, 4,5	11,2,6,7	2,11,1	11,2,1	2,7,11,1
1630	11,3,2,1, 4,5	11,2,6,7	2,11,1	11,2,1	2,7,11,1
1700	11,3,2,1, 4,5	11,2,6,7	2,11,1	11,2,1	2,7,11,1
1730	11,3,2,1, 4,5	11,2,6,7	2,11,1	11,2,1	2,7,11,1
1800	3,2,4,5	11,2,6,7	2,11,1	11,2,1	2,7,11,1
1830	3,2,4,5, 1	11,2,6,7	2,11,1	11,2,1	2,7,11,1 8
1900	3,2,4,5, 11,1	11,2,6,7, 1	2,11,1,3	11,2,1	2,7,11,1 8

key:

1. small scattered cumulus
2. cumulus
3. thin cirrus
4. cirrus
5. thin cirrus over cloud
6. stratus
7. stratocumulus
8. altocumulus
9. altostratus
10. cumulonimbus
11. clear
12. haze
13. fog

	7/18/91	7/19/91
1430	7,5,2,11, 1,12	8,2,7,12
1500	7,5,2,11, 1,12	8,2,7,12
1530	5,12	8,2,7,12
1600	5,2,12	8,2,1,11, 12
1630	5,2,7,12	2,7,8,11, 12
1700	11,5,2,12	1,8,11,1, 12
1730	11,3,2,12	5,2,11,8, 7,12
1800	11,2,12,8	2,1,12
1830	11,2,8,12, 7	2,1,12
1900	11,2,8,12, 7	2,1,12

key:

1. small scattered cumulus
2. cumulus
3. thin cirrus
4. cirrus
5. thin cirrus over cloud
6. stratus
7. stratocumulus
8. altocumulus
9. altostratus
10. cumulonimbus
11. clear
12. haze
13. fog